

Supplementary Material

Immune transcriptomic analysis on COVID-19 patients with varying clinical presentations identifies severity markers

Abdallah Musa Abdallah^{1*}, Hifzur Rahman Ansari², Muhammad Shuaib³, Asmma Doudin⁴, Hadi M. Yassine⁵, Syed Shahab Ud Din Shah¹, Fayaz Ahmad Mir⁶, Mohamed M. Emara¹, Abdel-Naser Elzouki⁷, Farhan S. Cyprian^{1*}

¹College of Medicine, QU Health, Qatar University, Doha, Qatar; ²King Abdullah International Medical Research Center (KAIMRC), King Saud Bin Abdulaziz University for Health Sciences (KSAU-HS), Ministry of National Guard-Health Affairs (MNG-HA), Jeddah, Saudi Arabia; ³Pathogen Genomics Laboratory, Bioscience Program, Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia; ⁴Laboratory of Immunoregulation, Research Department, Sidra Medicine, Doha, Qatar; ⁵Biomedical Research Center (BRC), Qatar University, Doha, Qatar, ⁶Qatar Metabolic Institute, Academic Health System, Hamad Medical Corporation, PO Box 3050, Doha, Qatar. ⁷Internal Medicine Department, Hamad General Hospital, Hamad Medical Corporation, Doha P.O. Box 3050, Qatar.

Correspondence*: Abdallah Musa Abdallah and Farhan S. Cyprian

Abdallah.musa@qu.edu.qa

fcyprian@qu.edu.qa

1 Supplementary Figures and Tables

1.1 Supplementary Figures

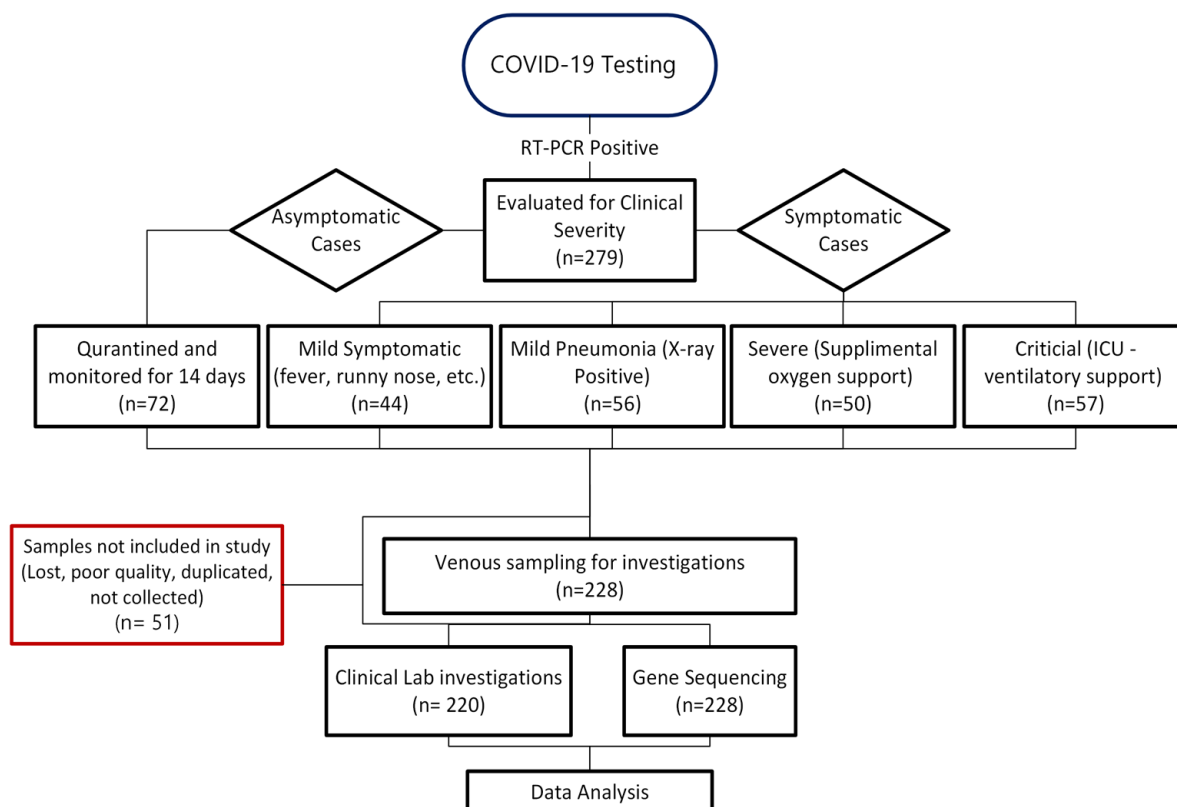


Figure S1: Selection and randomization process of patient samples that tested positive for SARS-CoV-2 by RT-PCR.

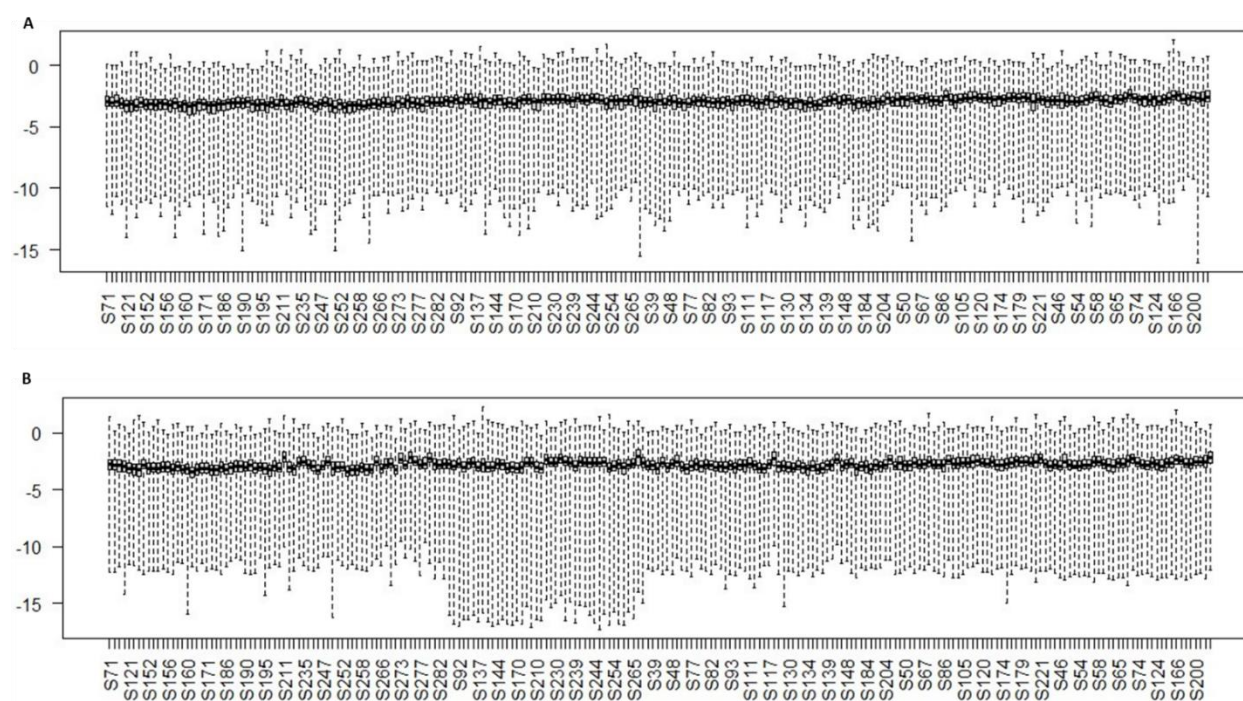


Figure S2: RNA-seq data intensity distribution before and after normalization. A. Raw and B. normalized RNA-seq data obtained for 228 COVID-19 patients. Dashed lines represent individual patients.

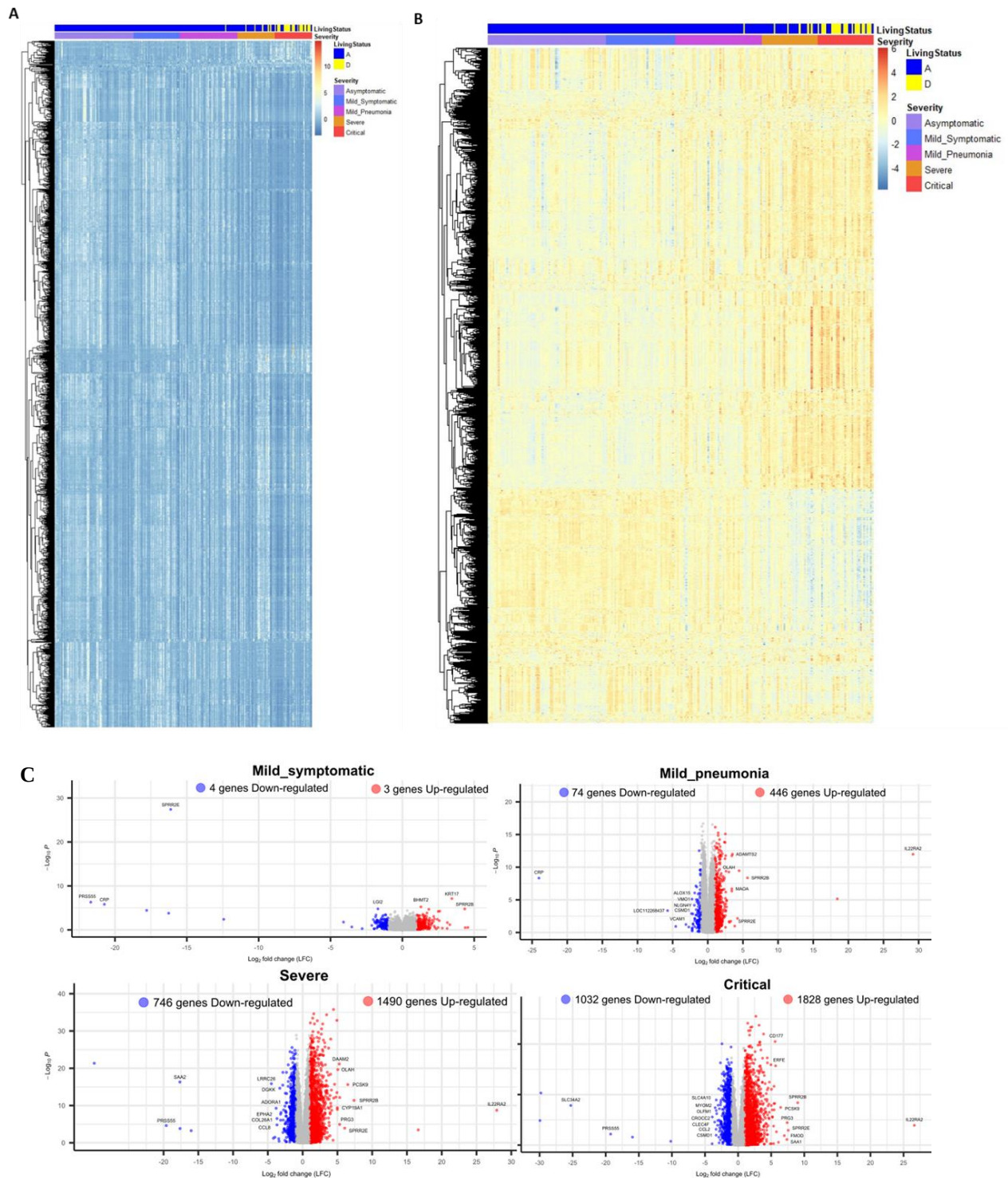


Figure S3: Hierarchical cluster analysis (HCA) of all samples sequenced by RNA-seq and genes filtered based on high variance (R library, "genefilter"). **A.** Heatmap reporting clustering of all samples and genes using Z-score of raw counts. **B.** Heatmap of selected 1932 genes filtered based on high variance (R library, "genefilter") after Variance stabilized normalization and Z-score. **C.** Volcano plot showing fold change for each gene and their significance level. Colored dots show genes with an adjusted p-value <0.05 and an absolute $\log_2(\text{fold change}) > 1$. Blue dots represent down-regulated genes, and red dots represent up-regulated genes. Numbers of over-expressed and under-expressed genes are indicated.

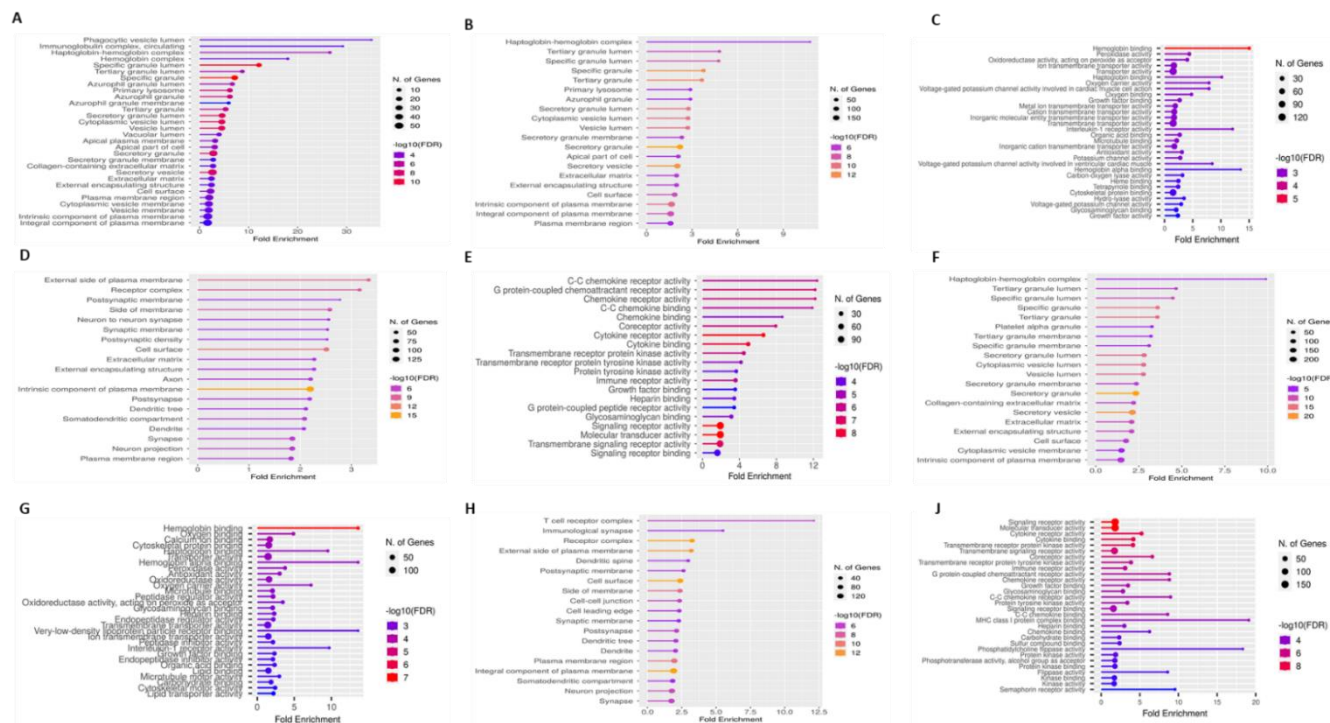


Figure S4. GO term: **A.** GO Cellular Component (CC): Mild Pneumonia up. **B.** GO CC: Severe up. **C.** GO Molecular Function (MF): Severe up. **D.** GO CC: Severe down. **E.** GO MF: Severe down. **F.** GO CC: Critical up. **G.** GO MF: Critical up. **H.** GO CC: Critical down. **J.** GO MF: Critical down.

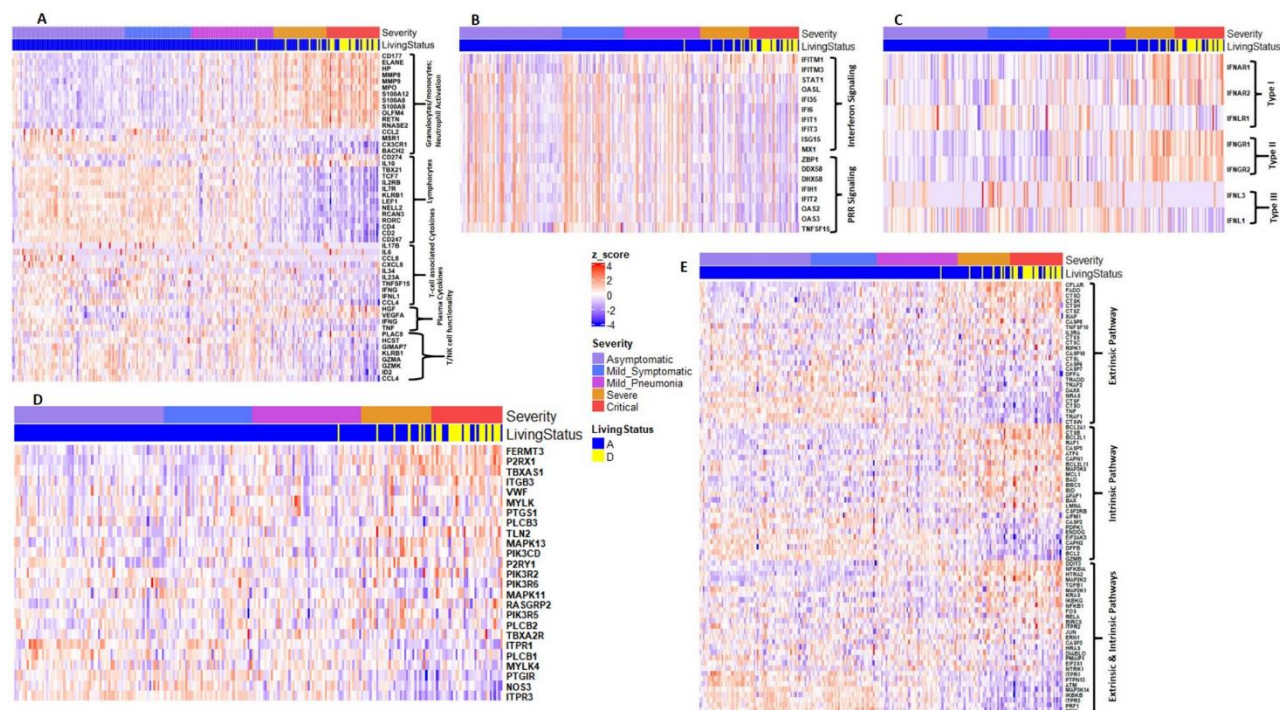


Figure S5: The expression level of multiple genes within several immunity categories and Apoptosis-related pathway in COVID-19 patients. Log₂ fold change value of multiple differentially expressed genes within several immunity categories. Heatmap of DEGs related to genes associated with **A.** granulocytes, monocytes, lymphocytes, cytokines and T/NK cell functionality; **B.** IFN and PRR signaling; **(C).** IFN; **D.** Platelets activation; **E.** Apoptosis-related signal pathway. Heatmaps of the DEGs, scaled based on log₂FC values, with blue and red colors indicating low and high expressions, respectively.

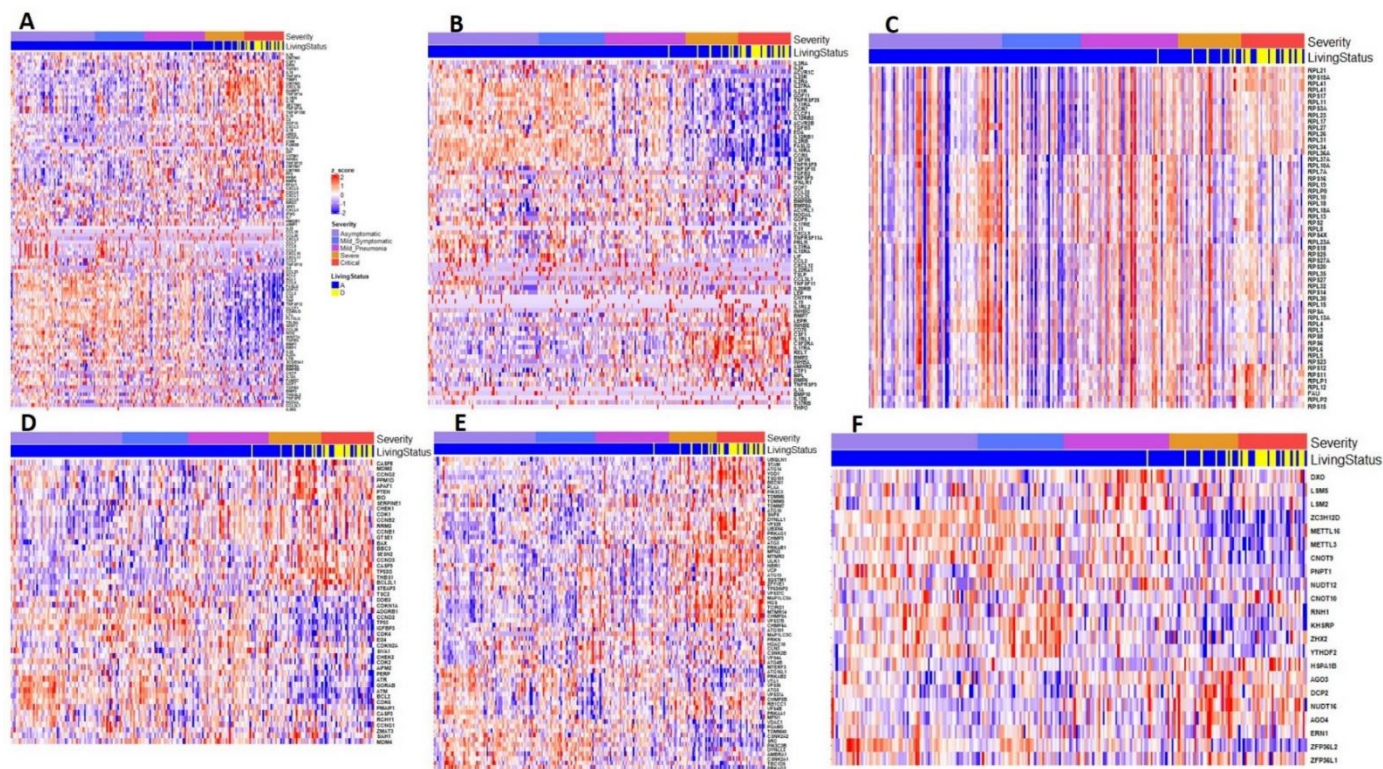
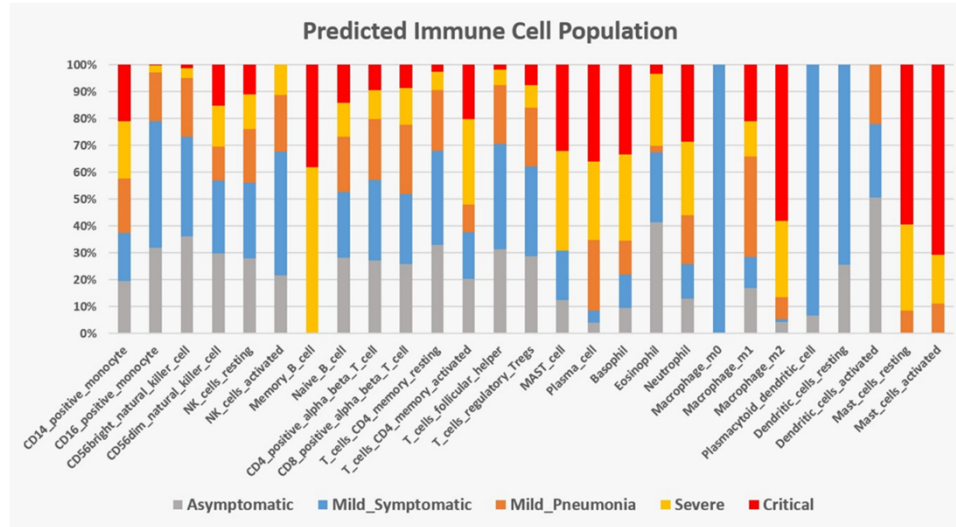


Figure S6: The expression level of multiple genes within several immunity categories and Apoptosis-related pathway in COVID-19 patients. Heatmap of DEGs related to genes associated with **A. Inflammatory Cytokines**. **B. Cytokine - cytokine Receptor** **C. Ribosome** **D. p53 signaling pathway** **E. Macroautophagy**; and **F. mRNA Catabolism**

A



B

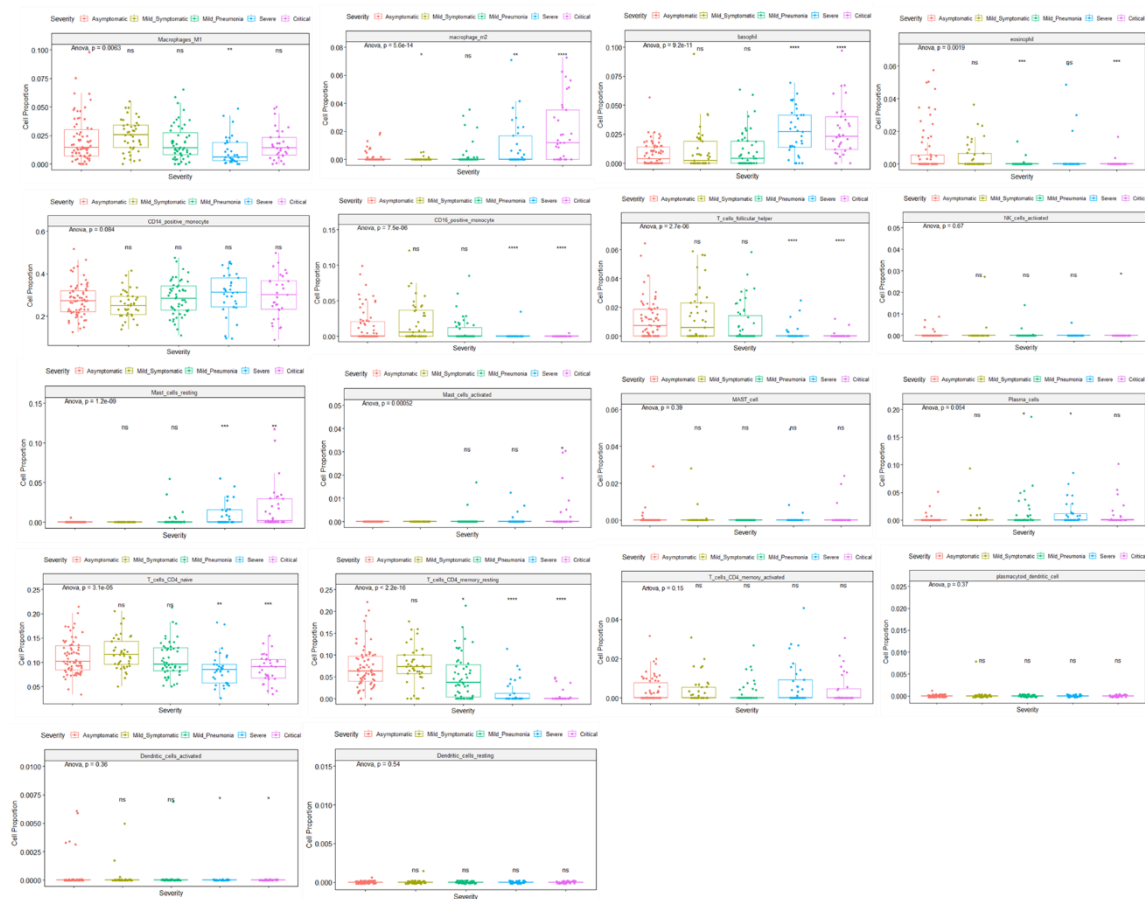


Figure S7: Immune cell composition predicted from transcriptome data. A. The average proportion of immune cell types predicted from transcriptome data using a deconvolution algorithm B. Proportions of blood cells were estimated from RNA sequencing data. Asterisks show significant difference (one-way anova test, Significance codes: >0.05 'ns', <0.05 '*', <0.01 '**', <0.001 '***', 0 '****'). Points represent individual patient data. p-values were calculated using a two-tailed Wilcoxon test.

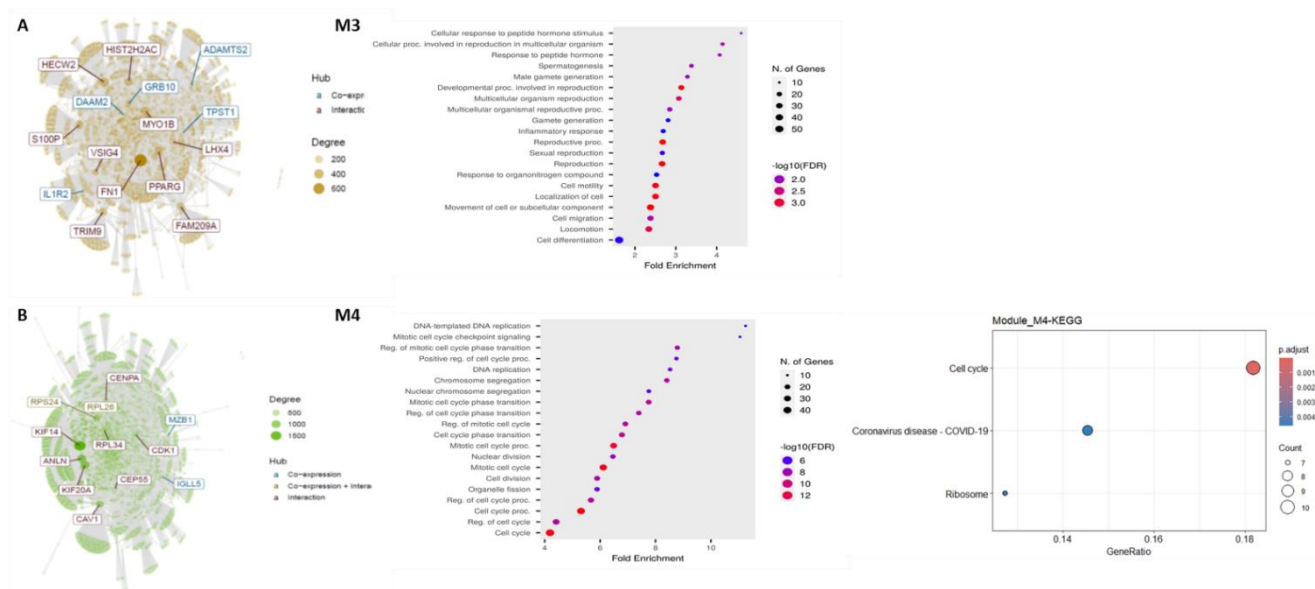


Figure S8: Co-expression gene modules present in blood transcriptomes of COVID-19 patients. A. Interaction plot for M3 and **B.**, which contains genes enriching different immunity-associated pathways, exhibited by the GO BP and KEGG pathway bar plot at the right side of figures **A** and **B**.

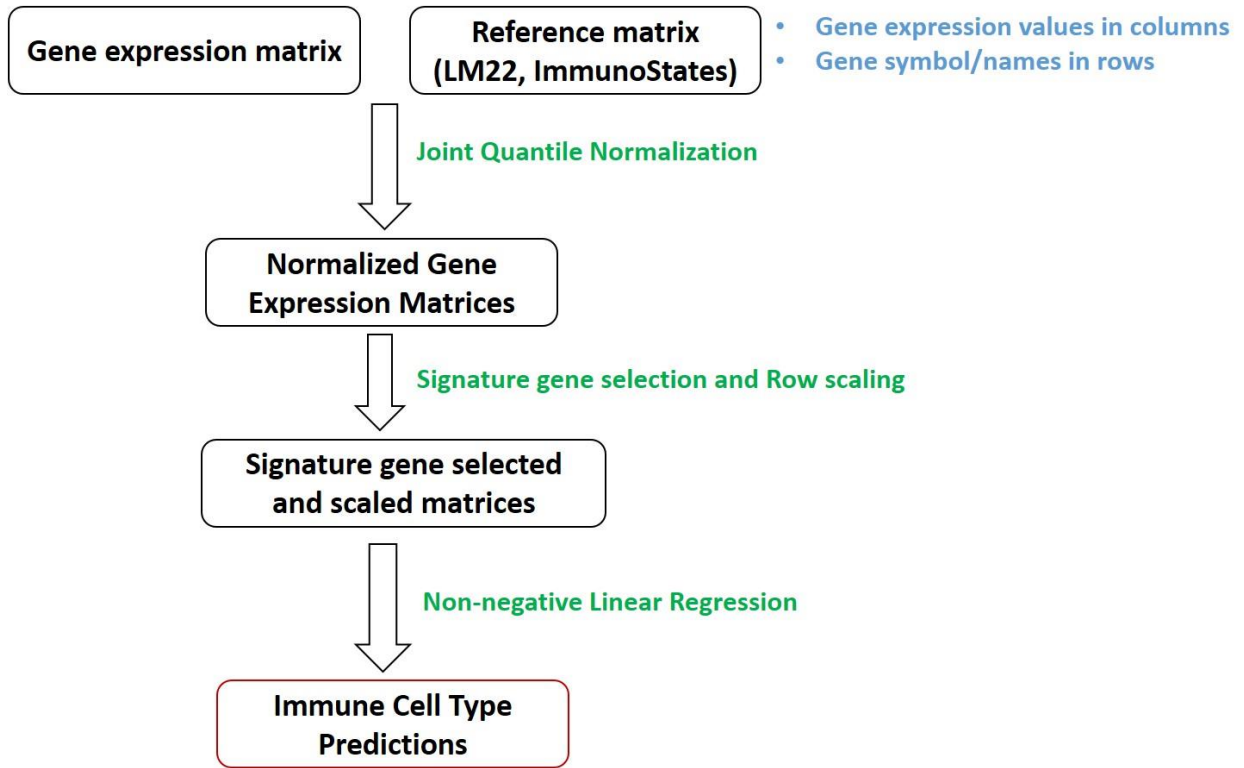


Figure S9: The GEDIT algorithm flowchart. Two input matrices are used (COVID 19 severity gene expression and reference LM22 + ImmunoStates). Both matrices are normalized through joint quantile normalization. Then only signature genes are selected and row scaled to remove the bias of highly or dominant expressed genes. Simple linear regression is used for the final prediction of immune cell type abundance.

1.2 Supplementary Tables

Table S1: Demographic characteristics of study participants categorized by stratified by asymptomatic, mild symptomatic, mild pneumonia, severe and critical Covid-19 cases.

Table S2: List of all differentially expressed genes without any filter.

Table S3: List of all differentially expressed genes filtered p value ≤ 0.05 .

Table S4: List of up and down regulated genes.

Table S5: List of common and unique genes.

Table 6: List of co-expression-module genes.

Table 7: Demographic table